

# A LINEAR THRESHOLD FOR UNIQUENESS OF SOLUTIONS TO RANDOM JIGSAW PUZZLES

ANDERS MARTINSSON

**ABSTRACT.** We consider a problem introduced by Mossel and Ross [*Shotgun assembly of labeled graphs*, arXiv:1504.07682]. Suppose a random  $n \times n$  jigsaw puzzle is constructed by independently and uniformly choosing the shape of each “jig” from  $q$  possibilities. We are given the shuffled pieces. Then, depending on  $q$ , what is the probability that we can reassemble the puzzle uniquely? We say that two solutions of a puzzle are similar if they only differ by permutation of duplicate pieces, and rotation of rotationally symmetric pieces. In this paper, we show that, with high probability, such a puzzle has at least two non-similar solutions when  $2 \leq q \leq \frac{2}{\sqrt{\epsilon}}n$ , all solutions are similar when  $q \geq (2 + \epsilon)n$ , and the solution is unique when  $q = \omega(n)$ .

## 1. INTRODUCTION

A *jigsaw puzzle* is a collection of square pieces where each of the four edges of a piece has a shape, referred to as a *jig*, so that it fits together with a subset of the edges of the other pieces. An *edge-matching puzzle* is a collection of square pieces where each side of every piece is given a color. The goal of the respective puzzles is to assemble the pieces into a certain form, in this paper this will always be an  $n \times n$  square, such that all pairs of adjacent pieces fit together. In the case of a jigsaw puzzle, this means that the jigs of edges that are aligned next to each other should have complementary shapes, and for an edge-matching puzzle, such pairs of edges should have the same color. Here we assume that the pieces are allowed to be rotated, but not flipped upside down.

In order to make this a bit more formal, we assume that there are  $q$  possible types of jigs, enumerated from 1 to  $q$ . A jig type can either be symmetric, so that it fits together with itself, or be part of a pair of complementary types that fit together with each other. We can describe this relation by fixing a map  $\iota$  from  $\{1, 2, \dots, q\}$  to itself such that  $\iota \circ \iota = id$  and a jig of type  $j$  fits with jigs of type  $\iota(j)$ . Note that we can consider an edge-matching puzzle as a special case of a jigsaw puzzle by taking  $\iota$  equal to the identity map.

In a recent paper by Mossel and Ross [5], a simple model for random edge-matching puzzles, later generalized to random jigsaw puzzles in [1], was proposed. We imagine that we start with an  $n \times n$  grid of identical unit squares. For each of the four sides of each piece, we choose its color, or type of jig respectively, out of  $q$  possibilities, under the restriction that pairs of connected sides must get the same color/complementary jig types. Note that this means that, unlike most real jigsaw puzzles, also edges along the boundary are assigned colors/jigs. We will refer to such an assignment of

colors to a puzzle as a *coloring*, and an assignment of jig types as a *carving*. The authors asked, suppose we chose one such puzzle uniformly at random, then what is the probability that the puzzle can be uniquely recovered from the shuffled collection of pieces? Further, how can this recovery be done efficiently? They called this problem shotgun assembly of a random jigsaw puzzle due to its similarity to genetic shotgun sequencing, which is a technique for sequencing a long DNA strand by sampling random subsequences.

The notion of “unique recovery” needs some elaboration. We consider a solution of the puzzle to be a positioning and orientation of the pieces into a fixed  $n \times n$  grid such that all adjacent pieces fit together. The solution consisting of all original positions and orientations will be referred to as the *planted assembly*. We here assume that, besides the choice of jigs/colors, all pieces are identical and rotationally symmetric. Hence any solution to the puzzle is equally likely to be the planted assembly. As the pieces can be rotated, the best we could ever hope for is to be able to recover the planted assembly up to a global rotation of the puzzle. Besides that, the puzzle may contain duplicate pieces, that is, two pieces with identical colors/jigs, or rotationally symmetric pieces, that is, opposite sides have the same colors/jig types. Note that such pieces automatically give rise to additional, albeit not very different, solutions to the puzzle.

We say that two solutions of a puzzle are *similar* if they only differ by a global rotation, permutation of duplicate pieces, and rotation of rotationally symmetric pieces, or, equivalently, if the solutions have the same coloring/carving up to global rotation. In the terminology of [1], a puzzle has *unique vertex assembly* (UVA) if the only solutions are the four global rotations of the planted assembly, and a puzzle has *unique edge assembly* (UEA) if all solutions are similar.

It was shown by Mossel and Ross [5] that, with high probability, a random edge-matching puzzle as above has at least two non-similar solutions when  $2 \leq q = o(n^{2/3})$ , and has a unique solution up to global rotation when  $q = \omega(n^2)$ . Recently, two papers, Bordenave, Feige and Mossel [1], and Nenadov, Pfister and Steger [6], considering this problem were published on arxiv.org, both on May 11:th 2016, and both proving essentially the same result: for  $q \geq n^{1+\varepsilon}$ , a random edge-matching puzzle has a unique solution up to global rotation with high probability for any fixed  $\varepsilon > 0$ , and for  $q = o(n)$  there are duplicate pieces with high probability and hence multiple, but possibly all similar, solutions. It should be noted that the paper by Bordenave et al. assumes that the pieces are not allowed to be rotated, but remarks in the last section of the paper how their argument can be modified slightly, both to allow rotations and to generalize to the random jigsaw puzzle model above.

Concerning the problem of how to recover the planted assembly efficiently, Bordenave et al. describe an algorithm that recovers it with high probability when  $q \geq n^{1+\varepsilon}$  with time complexity  $n^{O(1/\varepsilon)}$ . As a comparison, the general problems of finding one solution to a given  $n \times n$  jigsaw puzzle or edge-matching puzzle are known to be NP-complete [3], see also [2]. The problem also seems to be hard in practice. In the summer of 2007, a famous edge-matching puzzle, Eternity II, was released, with a \$2 million prize for the first

complete solution [7]. This puzzle consists of 256 square pieces that should be assembled into a  $16 \times 16$  square. There are in total 22 edge colors, not including the boundary, which is marked in gray. The competition ended on 31 December 2010, with no solution being found, and at the time of writing, the puzzle is claimed to remain unsolved.

The aim of this paper is to prove the following result regarding uniqueness of the solution of a random jigsaw or edge-matching puzzle.

**Theorem 1.1.** *As  $n \rightarrow \infty$  the following holds with high probability for a random jigsaw puzzle with  $q$  types of jigs or random edge-matching puzzle with  $q$  colors.*

- (i) *For  $2 \leq q \leq \frac{2}{\sqrt{e}}n$ , there are at least two non-similar solutions.*
- (ii) *For  $q \geq (2 + \varepsilon)n$ , for any fixed  $\varepsilon > 0$ , all solutions are similar.*
- (iii) *For  $q = \omega(n)$ , the solution is unique up to global rotation.*

We remark that a weaker form of (i) was first proved in a earlier version of this paper [4]. This will be shown again in this paper, but using a significantly simpler argument.

The question remains what happens in the interval  $\frac{2}{\sqrt{e}}n \leq q \leq 2n$ . To get a qualitative understanding for this range, we can compare our random model to a collection of  $n^2$  independently colored/carved pieces where each color/jig type is chosen uniformly at random, that is, without a planted solution. There are  $4^{n^2}(n^2)!$  ways to place and orient the pieces into an  $n \times n$  grid, and the probability that the pieces fit together in a given configuration is  $q^{-2n(n-1)}$ . Hence the expected number of solutions of such a puzzle is

$$\frac{4^{n^2}(n^2)!}{q^{2n(n-1)}} \approx \left( \frac{2n}{\sqrt{eq}} \right)^{2n^2}.$$

We see that there is a transition at  $q = \frac{2}{\sqrt{e}}n$  where the expected number of solutions goes from being very large to very small.

Connecting this back to our model, the fact that we force the puzzle have at least one solution may increase the probability that other ways to assemble the pieces are also solutions. On the other hand, based on the proofs in this article, my intuition is that typical solutions are either similar to the planted assembly, or have very little similarity to it, and hence this effect should be small. Because of this, I conjecture that the event that all solutions to a random jigsaw or edge-matching puzzle are similar has a sharp threshold at  $\frac{2}{\sqrt{e}}n + o(n)$ . Moreover, considering how strongly the estimates for  $\mathbb{P}(UEA)$  in Section 2 depend on  $q$ , I believe that this threshold is very sharp. It might even jump directly from  $o(1)$  to  $1 - o(1)$  when increasing the number of colors/jig types by 1.

Using a similar heuristic, we can explain the discrepancy between the bounds in parts (i) and (ii) of Theorem 1.1 (the factor  $\sqrt{e}$ ). In the proof of (i) we consider the entire assembled puzzle, so, heuristically, random solutions should stop affecting the calculations at  $q = \frac{2}{\sqrt{e}}n$ . On the other hand, the proof of (ii) is based on considering local solutions to the puzzle. The expected number of ways to build, say, a  $k \times k$  square of matching pieces out of  $n^2$  independently chosen pieces for  $k = o(n)$  is roughly  $\frac{4^{k^2}(n^2)^{k^2}}{q^{2k(k-1)}} \approx$

$\left(\frac{2n}{q}\right)^{2k^2}$ . Hence, in this case, solutions unrelated to the planted assembly should stop affecting the calculations only at  $q = 2n$ . Thus, it appears that new ideas are needed to close this gap in the main result.

The remainder of the paper will be structured as follows. In Section 2 we give a proof of part (i) of Theorem 1.1. Section 3 briefly investigates the probability of duplicates and rotationally symmetric pieces, which shows that (ii)  $\implies$  (iii). Finally, in Section 4 we prove part (ii). These sections will be formulated in terms of the random jigsaw puzzle model, but as we already noted, the random edge-matching puzzle can be considered as a special case of this.

## 2. PROOF OF THEOREM 1.1, PART (i)

Let us refer to the unordered collection of jigsaw pieces of a puzzle as the *box* of the puzzle. That is, the box contains the information of how many pieces of each combination of jig types there are in the puzzle, but no information beyond that about their locations or orientations in the planted assembly. The key observation of this proof is that any carving that results in a puzzle with UEA is uniquely determined, up to global rotation, by the box of the puzzle. Hence, there are at most 4 times more such carvings than there are possible boxes.

To make use of this observation, we need the following estimates. First, the probability that the planted assembly is given any specific carving is  $q^{-2n(n+1)}$ . Second, there are  $q^4$  ways to choose the jigs of a jigsaw piece:  $q$  of which being invariant under a  $90^\circ$  rotation,  $q^2 - q$  being invariant under a  $180^\circ$  but not a  $90^\circ$  rotation, and the remaining  $q^4 - q^2$  having no rotational symmetry. Hence there are  $\frac{q^4 - q^2}{4} + \frac{q^2 - q}{2} + q = \frac{q^4 + q^2 + 2q}{4}$  possible types of jigsaw pieces. From this it follows that the total number of boxes (including those without solutions) is  $\binom{\frac{1}{4}(q^4 + q^2 + 2q) + n^2 - 1}{n^2}$ .

Now, letting  $N_C$  denote the number carvings that result in a puzzle with UEA, we have

$$\begin{aligned} \mathbb{P}(UEA) &= q^{-2n(n+1)} \cdot N_C \leq q^{-2n(n+1)} \cdot 4 \binom{\frac{1}{4}(q^4 + q^2 + 2q) + n^2 - 1}{n^2} \\ &\leq 4q^{-2n(n+1)} \frac{\left(\frac{1}{4}(q^4 + q^2 + 2q) + n^2 - 1\right)^{n^2}}{n^2!} \\ &= \Theta\left(\frac{1}{n}\right) q^{-2n} \left(\frac{e(q^4 + q^2 + 2q + 4n^2 - 4)}{4q^2n^2}\right)^{n^2}, \end{aligned}$$

where we used Stirling's formula in the last step. One can observe that

$$\frac{e(q^4 + q^2 + 2q + 4n^2 - 4)}{4q^2n^2}$$

is convex in  $q$ , equals  $\frac{e}{4} + O(\frac{1}{n^2})$  for  $q = 2$  and  $1 + O(\frac{1}{n^2})$  for  $q = \frac{2}{\sqrt{e}}n$ . Hence, for any  $2 \leq q \leq \frac{2}{\sqrt{e}}n$ ,

$$\mathbb{P}(UEA) \leq \Theta\left(\frac{1}{n}\right) q^{-2n} \left(1 + O\left(\frac{1}{n^2}\right)\right)^{n^2} = \Theta\left(\frac{1}{n}\right) q^{-2n},$$

which tends to 0 as  $n \rightarrow \infty$ .  $\square$

### 3. PROOF OF THEOREM 1.1, (ii) IMPLIES (iii)

To understand the behavior of the probability that a random jigsaw puzzle has a unique solution, up to global rotation, it is natural to consider uniqueness of the solution as the intersection of two events: the event that all solutions of the puzzle are similar, and the event that the puzzle does not contain duplicate or rotationally symmetric pieces. The former is characterized by parts (i) and (ii) of Theorem 1.1, and, as stated in the introduction, I conjecture that it has a very sharp threshold at  $q = \frac{2}{\sqrt{e}}n + o(n)$ . It remains to consider the latter event.

**Proposition 3.1.** *The probability that a random jigsaw puzzle contains either duplicate or rotationally symmetric pieces is  $o(1)$  for  $q = \omega(n)$ .*

*Proof.* Let  $X$  denote the number of pairs of duplicate jigsaw pieces, and  $Y$  the number of rotationally symmetric pieces respectively. The probability that two given jigsaw pieces have identical jig types is  $\Theta(q^{-4})$  if the pieces are non-adjacent in the planted assembly, and  $O(q^{-3})$  if they are adjacent. Hence,

$$\mathbb{E}X = \Theta(n^4 q^{-4}) + O(n^2 q^{-3}).$$

Similarly, the probability that a jigsaw piece has rotational symmetry is  $q^{-2}$ , which implies that

$$\mathbb{E}Y = \Theta(n^2 q^{-2}).$$

By Markov's inequality it follows that

$$\mathbb{P}(X + Y \neq 0) \leq \Theta(n^4 q^{-4}) + O(n^2 q^{-3}) + \Theta(n^2 q^{-2}),$$

which tends to 0 for  $q = \omega(n)$ .  $\square$

By part (ii) of Theorem 1.1, all solutions of a random jigsaw puzzle are similar with high probability when  $q = \omega(n)$ . Part (iii) of Theorem 1.1 follows by combining this with Proposition 3.1.  $\square$

**Remark 3.2.** Considering the estimates for  $\mathbb{E}X$  and  $\mathbb{E}Y$  further, one would expect the probability of  $X = Y = 0$  to be bounded away from 0 and 1 for  $q = \Theta(n)$ . In particular, this would mean that the probability that a random jigsaw puzzle has a unique solution, up to global rotation, is bounded away from 0 and 1 when  $(2 + \varepsilon)n \leq q = O(n)$ . For the sake of brevity, we will not attempt to prove this here. We can however note that a partial result to this effect was shown in Section 3 of [6], namely that

$$\mathbb{P}(X = 0) \leq \exp\left(-\frac{n^4 - 2n^2}{8q^4}\right),$$

which implies that the probability of a unique solution is bounded away from 1 for  $q = O(n)$ . In particular, the bound  $q = \omega(n)$  in part (iii) of Theorem 1.1 is sharp.

## 4. PROOF OF THEOREM 1.1, PART (ii)

To simplify our argument below, let us consider a problem related to unique edge assembly – reconstruction. Given a box of shuffled pieces from a random jigsaw puzzle, we try to guess the planted assembly, up to similarity. Clearly, when all solutions are similar, any solution is valid. Hence it is possible to reconstruct the planted assembly correctly with probability at least  $\mathbb{P}(UEA)$ . On the other hand, if we condition on the event that we are given a certain box with at least two non-similar solutions and a corresponding guess, then there are at most four carvings of the planted assembly for which the guess will be valid, and at least one for which it will be invalid. As all carvings are equally likely, the guess is invalid with probability at least  $\frac{1}{5}$ . Hence any reconstruction tactic will fail with probability at least  $\frac{1}{5}(1 - \mathbb{P}(UEA))$ .

Our proof strategy can be summarized as follows: Throughout this section we will assume that  $q \geq (2 + \varepsilon)n$  for some fixed  $\varepsilon > 0$ . We introduce the notion of a feasible assembly being *k-good*. Here it is important that one can determine *k-goodness* without knowledge of the planted assembly. We prove that for  $\omega(\ln n) = k = o(n^{1/12})$ , the planted assembly is *k-good* with high probability, and any *k-good* feasible assembly must be similar to the planted one. This gives us a simple reconstruction strategy that succeeds with high probability – guess any *k-good* planted assembly, if such exist. By the above reasoning, this implies unique edge assembly with high probability, as desired.

Let us start by defining some concepts. We will consider a jigsaw piece to be a vertex with four cyclically ordered *half-edges*, representing the sides of the piece. A *complete assembly* of an  $n \times n$  puzzle is a positioning and orientation of the jigsaw pieces into an  $n \times n$  grid, regardless of whether the pieces fit together in this configuration. We will formally consider this as a bijective map from  $\{1, \dots, n\}^2$  to itself together with a map from  $\{1, \dots, n\}^2 \rightarrow \mathbb{Z}_4$ , interpreted as the position of and orientation of each piece relative to the planted assembly. Similarly, a *partial assembly* is a positioning and orientation of a subset of pieces in the jigsaw puzzle into a square grid. One important example of a partial assembly is a  $k \times k$  *window* obtained by picking an appropriate square from a complete assembly. For a given (complete or partial) assembly, we say that two half-edges are *connected* if they correspond to sides of two different jigsaw pieces that are aligned next to each other in the assembly. Note that any assembly can be considered as a graph by joining connected half-edges into edges. We say that an assembly is *feasible* if all connected pairs of half-edges have complementary jig types, that is, everything fits together in the assembly.

For any assembly, we have a natural notion of a dual graph. Considering the assembly geometrically in the plane, the vertices in this graph are the common corners of at least two jigsaw pieces, and the edges are the common sides of two pieces. Hence the edges of the dual graph correspond to the connected pairs of half-edges in the assembly. See Figure 1.

The *contour graph* is the induced subgraph of the dual graph consisting of all edges whose corresponding pairs of half-edges are not connected in the planted assembly, see Figure 2. We will refer to the connected components

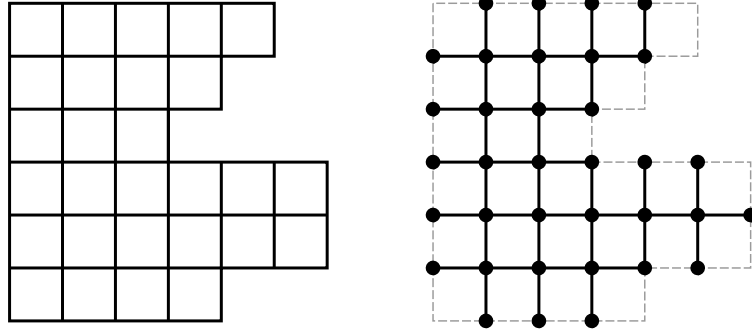


FIGURE 1. Illustration of a partial assembly (left) and the corresponding dual graph (right).



FIGURE 2. An example of a contour graph. Though not part of the random model, the puzzle is here given a motif to better visualize the relative position of the jigsaw pieces. The left picture illustrates the planted assembly of a  $5 \times 5$  puzzle, and the right one an alternative assembly. Edges along the contour graph are marked by bold line segments. Here, the contour graph contains three contours, and partitions the puzzle into 8 connected regions.

of the contour graph as *contours*. Note that the contour graph, together with the boundary of the assembly, partitions the pieces in the assembly into *connected regions* given by the sets of pieces contained in each face. An important observation is that, within each of these regions, the positions and orientations of pieces differ from the planted assembly by a common translation and rotation, and edges of the dual graph lies in the contour graph if and only if its half-edges come from different regions.

Given a feasible partial assembly, we say that an edge in its dual graph has *type*  $\{j, \iota(j)\}$  its corresponding half-edges have jig shape  $j$  and  $\iota(j)$  respectively. Let  $\mathcal{J} = \{\{j, \iota(j)\} : j = 1, 2, \dots, q\}$  be the set of between  $\frac{q}{2}$  and  $q$  possible types of edges in the dual graph. For a feasible assembly, and a set  $A$  of edges in its dual graph, the *shape multiplicity* of  $A$  is defined as

$$sm(A) = \sum_{J \in \mathcal{J}} \lfloor \#\{\text{edges in } A \text{ of type } J\} / 2 \rfloor.$$

Similarly, the shape multiplicity of a feasible partial assembly is the shape multiplicity of the set of all edges in its dual graph. Note that shape multiplicity can be interpreted as the maximal number of disjoint pairs of edges of the same type in  $A$ .

We say that a feasible complete assembly is *k-good* if the shape multiplicity of any  $k \times k$  window is at most 1 when the window touches the boundary of the puzzle, and at most 2 otherwise.

**Proposition 4.1.** *For  $k = o(n^{1/12})$ , the planted assembly is  $k$ -good with high probability.*

*Proof.* In order for a  $k \times k$  window of the planted assembly to have shape multiplicity at least  $i+1$ , there must exist  $i+1$  disjoint pairs of edges in the dual graph where each pair has a common type. By the union bound, the probability that this occurs in a carving of a  $k \times k$  window is  $O(k^{4(i+1)}q^{-(i+1)})$ . Summing this over  $O(n)$  windows with  $i = 1$  and  $O(n^2)$  ones with  $i = 2$  yields  $O(n k^{2.4}q^{-2} + n^2 k^{2.6}q^{-3}) = o(1)$ .  $\square$

For a given partial assembly, we would like to estimate the probability that it is feasible. One natural guess would be  $q^{-E}$  where  $E$  is the number of edges in the contour graph. However, this is not generally true as events of the form pairs of half-edges fit together may not be independent. In fact, in the case of an edge-matching puzzle, that is,  $\iota = id$ , one can show that such events are always increasing. In order to proceed, we make use of two observations. First, in order for there to be dependency, some connected regions must be adjacent in the planted assembly. We can formally describe this as that there are half-edges  $h_1, h_2$  in the contour graph, not necessarily connected in the new assembly (in fact, they cannot be connected there), such that  $h_1$  and  $h_2$  are connected in the planted assembly. Second, in the cases where there is dependency, some edges on the contour will get the same type.

**Lemma 4.2.** *Fix a partial assembly, and let  $E$  denote the number of edges in its contour graph  $C$ . If no two half-edges contained in  $C$  are connected in the planted assembly, then*

$$\mathbb{P}(\text{assembly is feasible}) = q^{-E}.$$

Moreover, without this restriction we have that for any  $i \geq 0$ ,

$$\mathbb{P}(\text{assembly is feasible} \wedge sm(C) \leq i) \leq q^{-E+i}.$$

*Proof.* We construct the abstract graph  $G$  whose vertices are the half-edges contained in  $C$ . We connect a pair of vertices by an *old edge* if they are connected in the planted assembly, and by a *new edge* if they are connected



in the partial assembly in the statement of the lemma. We will refer to such pairs of vertices in  $G$  as *old pairs* and *new pairs* respectively. Note that each vertex in  $G$  is the end-point of exactly one new edge and at most one old edge, so  $G$  consists of paths and cycles. Further, these components alternate between old and new edges, and, in the case of a path, begin and end with new edges.

The way the carving of the planted assembly is chosen means that each vertex in  $G$  that is not part of an old pair is independently and uniformly assigned a jig type, and each old pair is independently and uniformly assigned a pair of complementary jig types. Further, the new assembly is feasible if the jigs of half-edges in each component alternate between two complementary types. We say that the component is *feasible* if this holds.

In the case where no two half-edges contained in the contour graph are connected in the planted assembly, there are no old edges in  $G$ . Hence each of the  $E$  new pairs fits together independently with probability  $\frac{1}{q}$ .

Considering the latter case, a path in  $G$  with  $E'$  new edges,

$$h_1 \rightarrow h_2 \rightarrow \cdots \rightarrow h_{2E'},$$

is feasible if  $h_2, h_4, \dots, h_{2E'}$  all have the complementary jig type to  $h_1$ , which occurs with probability  $q^{-E'}$ . Similarly, a cycle in  $G$  with  $E'$  new edges,

$$h_1 \rightarrow h_2 \rightarrow \cdots \rightarrow h_{2E'} \rightarrow h_1,$$

is feasible if  $h_2, h_4, \dots, h_{2E'}$  all have the same jig type, which occurs with probability  $q^{-E'+1}$ . As the total number of new edges is  $E$ , we see that the probability that the new assembly is feasible is  $q^{-E}$  when  $G$  only consists of paths, and increases by a factor of  $q$  for each cycle.

Note that, by the definition of contour graph, a pair of vertices in  $G$  cannot be connected by both an old and a new edge. Hence, each cycle contains at least two new edges. As a consequence, for each feasible cycle in  $G$  there are two edges of the same type in  $C$ . Hence, if the partial assembly is feasible,  $sm(C)$  is at least the number of cycles in  $G$ . In particular, there are either more than  $i$  cycles, in which case the probability of the event in the statement is 0, or there are at most  $i$  cycles, in which case the probability is at most  $q^{-E+i}$ , as desired.  $\square$

We further need some way of counting the number of possible shapes of contours in assemblies. To this end, we have the following estimate.

**Lemma 4.3.** *Up to translation, the number of connected subgraphs  $G \subset \mathbb{Z}^2$  with  $E$  edges and  $F$  bounded faces is at most  $\binom{3E-4F+4}{2E-4F+4}$ . Moreover, for any  $\varepsilon' > 0$ , there exists a  $M = M_{\varepsilon'} > 0$  such that*

$$\binom{3E-4F+4}{2E-4F+4} = O_{\varepsilon'}(M^{E-2F}(1+\varepsilon')^E).$$

*Proof.* For any such  $G$ , let  $V$  denote its number of vertices. Then, by the Euler characteristic formula, we have

$$V - E + F = 1.$$

Note in particular that this means that  $V$  is independent of the choice of  $G$ . It further follows that

$$R := \sum_{v \in G} 4 - \deg(v) = 4V - 2E = 2E - 4F + 4.$$

Now, consider the following procedure for constructing  $G$ . Since we are only interested in the number of graphs up to translation, we may, without loss of generality, assume that the origin is a vertex in  $G$ . Let  $G_0$  be the empty graph, and let  $G_1$  be the graph only containing the origin. For each  $i \geq 1$ , we construct  $G_{i+1}$  by selecting a subset of the edges between  $V(G_i) \setminus V(G_{i-1})$  and  $\mathbb{Z}^2 \setminus V(G_i)$ , and adding them to  $G_i$ . If  $G_{i+1} = G_i$ , then we terminate and put  $G = G_i$ .

In order for the procedure to construct a graph with  $E$  edges dividing the plane into  $F$  regions, it must choose to include an edge  $E$  times, and choose not to do so  $R$  times. Hence we can encode any such graph as a binary string of  $E$  ones and  $R$  zeroes, which can be done in  $\binom{E+R}{R}$  ways.

Finally, considering the double sum

$$\sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \binom{i+j}{j} x^i y^j = \frac{1}{1-x-y},$$

which converges absolutely for  $|x| + |y| < 1$ , it follows that for any  $i, j \in \mathbb{N}_0$  and any  $x, y \geq 0$  such that  $x + y < 1$  we have

$$\binom{i+j}{j} \leq \frac{x^{-i} y^{-j}}{1-x-y} = O_{x,y}(x^{-i} y^{-j}).$$

Letting  $i = E$ ,  $j = 2E - 4F + 4$ ,  $x = (1 + \varepsilon')^{-1}$  and choosing  $M = M_{\varepsilon'}$  sufficiently large so that  $y = M^{-1/2} < 1 - x$ , we get

$$\binom{3E - 4F + 4}{2E - 4F + 4} = O_{\varepsilon'}((1 + \varepsilon')^E M^{E-2F+2}) = O_{\varepsilon'}((1 + \varepsilon')^E M^{E-2F}).$$

□

Now, suppose we pick an arbitrary  $k$ -good feasible complete assembly. If this assembly contains a contour with exactly four edges which surround a single jigsaw piece, then the eight pieces around it form the neighbourhood of an identical piece in the planted assembly. Hence by possibly moving around identical pieces, that is, exchanging the assembly for a similar one, we may assume that no such contours exist in the assembly.

We will now prove that if there is such an assembly with a non-empty contour graph, then this would imply the existence of certain feasible partial assemblies, which with high probability do not exist. Hence, all  $k$ -good feasible assemblies are similar to the planted assembly, as desired. We do this in three steps, depending on the type of contour.

**Case 1:** Contours not intersecting the boundary of the assembly, where the vertical and horizontal distances between pairs of vertices are each at most  $k - 2$ .

Suppose that the assembly contains such a contour  $C$ , with  $E$  edges and  $F$  bounded faces. Note that  $C$  can be contained in a  $k \times k$  window, hence

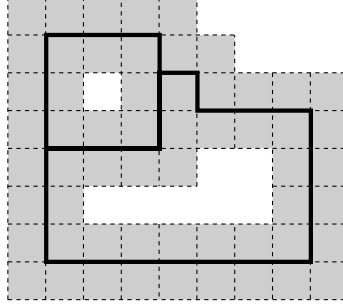


FIGURE 3. Example of a minimal partial assembly around a contour. Here, the contour has two bounded faces, partitioning the assembly into three connected regions.

$sm(C) \leq 2$ . By taking the subset of jigsaw pieces that have at least one corner on  $C$ , we obtain a feasible partial assembly with contour graph  $C$ ,  $F + 1$  connected regions, and satisfying  $sm(C) \leq 2$ . This is illustrated in Figure 3.

We now use Markov's inequality to show that, with high probability, there are no such assemblies in the random jigsaw puzzle. Consider a connected graph  $C \subset \mathbb{Z}^2$  with  $E$  edges and  $F$  bounded faces. Note that  $C$  uniquely defines a minimal surrounding configuration of jigsaw pieces. Without any restrictions on whether half-edges along the contour should be connected in the planted assembly, there are at most  $(4n^2)^{F+1}$  ways to choose the positions and orientations of the  $F + 1$  connected regions in the planted assembly. In counting the number of these where some pair of half-edges in  $C$  are connected in the planted assembly, we can first choose a pair of such half-edges. This relates the position and orientation of two connected regions, hence the positions and orientations of connected regions can be chosen in at most  $(4n^2)^F$  ways. By Lemma 4.2, the expected number of corresponding feasible assemblies is at most

$$(4n^2)^{F+1} q^{-E} + \binom{2E}{2} (4n^2)^F q^{-E+2} \leq O(E^2) (2n)^{2F-E+2} \left(1 + \frac{\varepsilon}{2}\right)^{-E}.$$

Note that  $C$  is contained in a  $k \times k$  window, hence  $E = O(k^2)$  and we can replace  $O(E^2)$  by  $O(k^4)$  in the right hand side.

Applying Lemma 4.3 with  $\varepsilon'$  chosen such that  $(1 + \varepsilon')(1 + \frac{\varepsilon}{2})^{-1} \leq (1 + \frac{\varepsilon}{3})^{-1}$ , the expected number of such assemblies is bounded by the sum

$$O_\varepsilon(n^2 k^4) \sum_E \left(1 + \frac{\varepsilon}{3}\right)^{-E} \sum_F \left(\frac{M}{2n}\right)^{E-2F},$$

where  $E$  and  $F$  goes over all possible combinations of numbers of edges and faces. Making the substitution  $T = E - 2F$  and letting  $T_0$  denote the minimal possible value of  $E - 2F$ , we can bound the sum by

$$O_\varepsilon(n^2 k^4) \sum_{E=0}^{\infty} \left(1 + \frac{\varepsilon}{3}\right)^{-E} \sum_{T=T_0}^{\infty} \left(\frac{M}{2n}\right)^T = O_\varepsilon(n^{2-T_0} k^4).$$

To estimate  $T_0$ , consider the sum of perimeters of all bounded faces of such a contour  $C$ . On the one hand, this will count every interior edge in  $C$  twice, and every edge on the boundary once, hence the sum equals  $2E - P$  where  $P$  is the perimeter of  $C$ . On the other hand, each face has perimeter at least four, hence the sum is at least  $4F$ . We conclude that  $E - 2F \geq \frac{P}{2}$ , and as we assumed that  $C$  surrounds more than just one jigsaw piece we must have  $E - 2F \geq \frac{P}{2} \geq 3$ . Hence  $T_0 \geq 3$ , as desired.

**Case 2:** No contour satisfies Case 1, but there is a contour with at least one vertex at (geometrical) distance  $\geq \frac{k}{2}$  from the boundary.

We assume that  $k$  is even. Suppose there is a  $k$ -good feasible assembly with a contour of this form. Consider a  $k \times k$  window centered at some vertex on this contour. Let  $C$  be the contour graph of the partial assembly consisting of this window, and let  $S$  be the surrounding square. Note that since the assembly is  $k$ -good,  $sm(C) \leq 2$ , and as no contour in the complete assembly satisfy Case 1, all contours of  $C$  must reach the boundary of the  $k \times k$  window. Hence if such a contour exists, then there is a connected graph  $C \cup S$  whose boundary  $S$  is a square of side length  $k$  and such that the midpoint of the square is a vertex in  $C$ , together with a feasible partial assembly in this square with contour graph  $C$  such that  $sm(C) \leq 2$ .

We again use Markov's inequality to show that this, with high probability, is impossible. Fix such a pair  $C, S$ , let  $E$  denote the number of edges of  $C$ , and  $F$  the number of bounded faces of  $C \cup S$ . By the same argument as in Case 1, the expected number of corresponding partial assemblies is at most

$$(4n^2)^F q^{-E} + \binom{2E}{2} (4n^2)^{F-1} q^{-E+2} = O(k^4)(2n)^{2F-E} \left(1 + \frac{\varepsilon}{2}\right)^{-E}.$$

Using Lemma 4.3 to count the number of possibilities for  $C \cup S$ , the probability that there exists a feasible partial assembly as above is at most

$$(1) \quad O_\varepsilon(k^4 M^{4k}) \sum_E \left(1 + \frac{\varepsilon}{3}\right)^{-E} \sum_F \left(\frac{M}{2n}\right)^{E-2F},$$

where, again,  $E$  and  $F$  runs over all possible combinations of numbers of edges and faces.

**Lemma 4.4.** *Let  $C, S, E$  and  $F$  be as above. Then  $E - 2F \geq \frac{k}{4} - 2 - O\left(\frac{E}{k}\right)$ .*

*Proof.* Enumerate the bounded faces of  $C \cup S$  from 1 to  $F$  and let  $P_i$  denote the perimeter of the  $i$ :th face. Then, as every edge in  $C$  is counted in the perimeter of two faces, and every edge in  $S$  in one face, we get

$$(2) \quad \sum_{i=1}^F P_i = 2E + 4k.$$

The perimeter of each face can be bounded in terms of its area. Suppose face  $i$  has area  $A_i$ , and that it contains jigsaw pieces from  $w_i$  columns and  $h_i$  rows. Then, on the one hand  $A \leq w_i \cdot h_i$ , and on the other hand the face contains two horizontal and vertical edges for each of these columns and rows. Hence

$$P_i \geq 2w_i + 2h_i \geq 2w_i + 2A_i/w_i \geq 4\sqrt{A_i}.$$

For  $A_i > \frac{15}{16}k^2$  we can improve this lower bound using the fact that the midpoint of the square is a vertex in  $C$ . Let us consider  $C \subseteq \mathbb{Z}^2$  such that the midpoint is  $(0, 0)$ , that is, the square is the area  $[-\frac{k}{2}, \frac{k}{2}]^2$ . Let  $\Gamma$  be a minimal length path from  $(0, 0)$  to the boundary of  $[-\frac{k}{4}, \frac{k}{4}]^2$ . We may, without loss of generality, assume that  $\Gamma$  hits the boundary on the line  $x = \frac{k}{4}$ . Note that this means that the  $y$ -coordinate of any vertex along  $\Gamma$  lies between  $-\frac{k}{4}$  and  $\frac{k}{4}$ . Then, in each column  $[l, l+1] \times \mathbb{R}$  for  $l = 0, 1, \dots, \lceil \frac{k}{4} \rceil - 1$  there is either at least 4 horizontal edges on the boundary of the face, or at most  $\frac{3k}{4}$  area units inside the face. Letting  $x$  be the number of columns satisfying the former, and  $y$  the number that satisfies the latter, we get

$$\begin{aligned} P_i &\geq 4\sqrt{A_i} + 2x, \\ A_i &\leq k^2 - \frac{k}{4}y, \\ x + y &\geq \frac{k}{4}. \end{aligned}$$

Hence,

$$P_i \geq 4\sqrt{A_i} + 2x \geq 4\sqrt{A_i} + \frac{k}{2} - 2y \geq 4\sqrt{A_i} + \frac{8}{k} \left( A_i - \frac{15}{16}k^2 \right).$$

Let  $f(a) = 4\sqrt{a} + \frac{8}{k} \max(0, a - \frac{15}{16}k^2)$ . This function is continuous, and concave on the intervals  $[0, \frac{15}{16}k^2]$  and  $[\frac{15}{16}k^2, \infty]$ . Then, by (2) and for a given  $F$ , the value of  $2E + 4k$  is bounded from below by the minimal value of  $\sum_{i=1}^F f(a_i)$  subject to  $\sum_{i=1}^F a_i = k^2$  and  $a_i \geq 1$  for  $i = 1, 2, \dots, F$ .

If  $F - 1 \geq \frac{1}{16}k^2$ , we have  $a_i \leq \frac{15}{16}k^2$  and hence the minimization problem is concave. In this case the function attains its minimum at an extreme point on the boundary of the domain. Up to permutation of variables, there is only one such point, namely  $a_1 = k^2 - (F - 1)$  and  $a_2 = a_3 = \dots = a_F = 1$ , yielding a minimum of  $4\sqrt{k^2 - (F - 1)} + 4(F - 1)$ .

On the other hand, if  $F - 1 < \frac{1}{16}k^2$ , we can divide the problem into two concave minimization problems by considering the case where all  $a_i$ 's are less than  $\frac{15}{16}k^2$  and the case where at least one variable, say  $a_1$ , is at least  $\frac{15}{16}k^2$ . Again, up to permutation of variables, the only extremal points are

$$\begin{aligned} a_1 &= \frac{15}{16}k^2, a_2 = \frac{1}{16}k^2 - (F - 2), a_3 = a_4 = \dots = a_F = 1, \\ a_1 &= k^2 - (F - 1), a_2 = a_3 = \dots = a_F = 1, \end{aligned}$$

with the corresponding values

$$4\sqrt{\frac{15}{16}k^2} + 4\sqrt{\frac{1}{16}k^2 - (F - 2)} + 4(F - 2) = (\sqrt{15} + 1)k + 4(F - 2) - O\left(\frac{F}{k}\right),$$

and

$$4\sqrt{k^2 - (F - 1)} + \frac{8}{k} \left( \frac{1}{16}k^2 - (F - 1) \right) + 4(F - 1) = \frac{9}{2}k + 4(F - 1) - O\left(\frac{F}{k}\right).$$

Hence, the minimum in this case is  $\frac{9}{2}k + 4(F - 1) - O(\frac{F}{k})$ .

Note that for  $F - 1 \geq \frac{1}{16}k^2$ , we have  $-\frac{9}{2}k = O(\frac{F}{k})$ . Hence, for all  $F$ , we can write the minimum as  $\frac{9}{2}k + 4(F - 1) + O(\frac{F}{k})$ .

In conclusion, we have  $2E + 4k \geq \frac{9}{2}k + 4(F - 1) + O(\frac{F}{k})$ . Hence  $E - 2F \geq \frac{1}{4}k - 2 + O(\frac{F}{k})$ , where clearly  $F = O(E)$ .  $\square$

Using Lemma 4.4, we can bound (1) by

$$O_\varepsilon(k^4 M^{(4+\frac{1}{4})k-2} n^{2-\frac{k}{4}}) \sum_{E=0}^{\infty} \left( \frac{(2n/M)^{O(\frac{1}{k})}}{1 + \frac{\varepsilon}{3}} \right)^E.$$

As  $k = \omega(\ln n)$ ,  $(2n/M)^{O(\frac{1}{k})} \rightarrow 1$  as  $n \rightarrow \infty$ . Hence the sum is at most  $O_\varepsilon(k^4 M^{(4+\frac{1}{4})k-2} n^{2-\frac{k}{4}}) \rightarrow 0$  as  $n \rightarrow \infty$ .

**Case 3:** All vertices in the contour graph have distance  $< \frac{k}{2}$  from the boundary.

If this occurs, there is a large connected region in the assembly that contains the  $(n - 2k) \times (n - 2k)$  square of all jigsaw pieces at distance at least  $k$  from the boundary. Consider the area in the planted assembly that corresponds to this square. In order to fit in the planted assembly, this area must cover all jigsaw pieces that are not in the  $2k$  outermost layers. As a consequence of this, any jigsaw piece in the  $k$  outermost layers of the new assembly must come from the  $2k$  outermost layers of the planted assembly.

Suppose we take a  $3 \times 3$  window in the assembly with a non-trivial contour graph  $C$ . Let again  $S$  be the boundary of the window,  $E$  the number of edges of  $C$ , and  $F$  the number of bounded faces of  $C \cup S$ . As this window is at distance  $\leq k$  from the boundary, it is part of a  $k \times k$  window that touches the boundary. Hence, by the definition of  $k$ -good,  $sm(C) \leq 1$ . Furthermore, by the above reasoning, all connected regions come from the  $2k$  outermost layers of the planted assembly.

As before we apply Markov's inequality to the number of such partial assemblies that can be constructed from the random jigsaw puzzle. There are at most  $4 \cdot 8kn$  ways to choose the orientation and position of each connected region in the planted assembly, hence, by Lemma 4.2, the expected number of such assemblies corresponding to a fixed  $C$  is at most

$$(32kn)^F q^{-E} + O(1)(32kn)^{F-1} q^{-E+1} = \left(1 + O\left(\frac{1}{k}\right)\right) (32kn)^F n^{-E}.$$

We note that if  $E > F$ , this is bounded by  $O(k^F n^{F-E}) = O(\frac{k^9}{n}) \rightarrow 0$  as  $n \rightarrow \infty$ . Hence, with high probability, only  $3 \times 3$  windows where  $E \leq F$  are possible.

Let us consider which non-trivial contour graphs  $C$  satisfy  $E \leq F$ . It is clear that the only possible way for  $C \cup S$  not to be connected is if  $C$  consists of four edges surrounding the center jigsaw piece, which we already assumed does not occur in our  $k$ -good feasible complete assembly (also,  $E = 4 > F = 2$  in this case). On the other hand, if  $C \cup S$  is connected, its Euler characteristic implies that  $E - F = V - 13$ , where  $V$  denotes the number of vertices of  $C \cup S$ . Note that there are always 12 vertices on the boundary of  $C \cup S$ , and, unless  $C$  is empty, it must contain at least one additional vertex. Hence, the only possibility is that  $C$  contains exactly one

interior vertex. It follows that  $C$  consists of two edges that separate a corner piece from the rest of the  $3 \times 3$  window.

Suppose we have a  $k$ -good feasible complete assembly where the contour graph of each  $3 \times 3$  window is either empty, or consists of two edges that cut out a corner piece. Observe that, in the latter case, if we can shift the window one step either vertically or horizontally towards the surrounded corner piece, then we would obtain a window with some other non-trivial contour graph. Hence, whenever a  $3 \times 3$  window has a non-trivial contour graph, we cannot make this shift. But the only reason why it would not be possible is if the surrounded corner is, in fact, a corner piece of the complete assembly.

In conclusion, for any such  $k$ -good feasible complete assembly, all jigsaw pieces except possibly the four corners have the correct relative positions. But, with high probability, the 16 sides of the corner pieces all have different jig shapes, and thus only fit together with the rest of the puzzle as in the planted assembly, as desired.  $\square$

## REFERENCES

- [1] Bordenave, C., Feige, U., and Mossel, E., *Shotgun assembly of random jigsaw puzzles* (2016+), available at <http://arxiv.org/abs/1605.03086>
- [2] Bosboom, J., Demaine, E.D., Demaine, M.L., Hesterberg, A., Manurangsi, P., and Yodpinyanee, A., *Even  $1 \times n$  Edge-Matching and Jigsaw Puzzles are Really Hard* (2016+), available at <http://arxiv.org/abs/1701.00146>
- [3] Demaine, E.D. and Demaine, M.L. *Jigsaw Puzzles, Edge Matching, and Polyomino Packing: Connections and Complexity*, Graphs and Combinatorics 23 (2007), Suppl 1, 195–208
- [4] Martinsson, A., *Shotgun edge assembly of random jigsaw puzzles* (2016+), available at <http://arxiv.org/abs/1605.07151>
- [5] Mossel, E. and Ross, N., *Shotgun assembly of labeled graphs* (2015+), available at <http://arxiv.org/abs/1504.07682>
- [6] Nenadov, R., Pfister, P., and Steger, A., *Unique reconstruction threshold for random jigsaw puzzles* (2016+), available at <http://arxiv.org/abs/1605.03043>
- [7] Description of Eternity II release, PR Newswire, 16 February 2007, <http://www.prnewswire.co.uk/news-releases/eternity-puzzle-is-back-this-summer-with-a-us2-million-prize-for-the-first-person-to-find-a-solution-153539245.html>

DEPARTMENT OF MATHEMATICAL SCIENCES, CHALMERS UNIVERSITY OF TECHNOLOGY AND UNIVERSITY OF GOTHENBURG, 41296 GOTHENBURG, SWEDEN  
*E-mail address:* `andemar@chalmers.se`